

Novel Big Data-supported dynamic toll charging system: Impact assessment on Portugal's shadow-toll highways

Abstract

Traffic congestion is a huge problem in many countries. It affects not only the inner workings of cities but also the quality of life of the people that endure it. In Portugal, traffic congestion happens mainly on national/urban roads, and this phenomenon has increased since the introduction of the so called shadow-toll systems in highways that were free to use. This work proposes a toll charging system that relies on a novel dynamic congestion charging scheme, supported by state of the art Big Data technologies, in order to shift traffic from national/urban roads to tolled highways, taking into account not only the Quality of Service of the highways and national roads, but also the competitiveness of toll prices for users. This Intelligent Transportation System was tested and validated in a real-world scenario with one of the biggest freight logistics companies in Portugal and with the Portuguese public road infrastructure operator.

Highlights

- A novel dynamic toll charging scheme applied to freight vehicles for Portugal's shadow-toll highways
- A Big Data platform supporting the Dynamic Toll Charging system
- Real-world test and validation of the Dynamic Toll Charging system, through the participation of a road operator and a freight logistics company

Keywords: Intelligent Transportation Systems, Big Data, Dynamic Toll Charging.

1. Introduction

Over the last two decades, the revolution led by the Internet and Information and Communications Technology (ICT) in general, has led to a profound change in the way we live, interact and work. The mobility and transportation sector has had a great impact in our daily lives, in particular in the way we commute, where the implementation of novel ICT approaches is contributing to a better quality of life. Nevertheless, worldwide population growth predominantly in cities and urban areas (United Nations, 2018), has resulted in new mobility-related challenges. Traffic congestions, resulting in more accidents and higher CO₂ emissions are some of the most problematic consequences of such a population growth. Expanding traffic network capacities by building more roads is extremely costly, as well as environmentally damaging. A more efficient usage of the existing network is vital in order to sustain the growing travel demand.

Intelligent Transportation Systems (ITS) is seen as plausible solution to tackle some of the challenges presented above. ITS is an emerging research area, which aims to tackle mobility and transportation challenges, while enabling personalized intelligent assistance to citizens in their daily journey. According to the EU, "ITS refer to the integration of ICT with vehicles and transport infrastructure to improve economic performance, safety, mobility and environmental sustainability for the benefit of all European citizens." (European Union, 2010). The implementation of ITS allows traffic and road infrastructure operators to have a better understanding of traffic conditions throughout the network, with the support of a variety Internet of Things (IoT) implementations. These include, sensors embedded within the transport infrastructure itself (e.g. road sensors, cameras and radio-frequency identification readers), mobile phones and social media feeds, generating a massive inflow of data that can then be used to generate new knowledge about the real-time

status of the infrastructure and predict about future conditions, supporting road operators in their decision-making processes (Guerreiro, Figueiras, Silva, Costa, & Jardim-Gonçalves, 2016).

Hence, to acquire such knowledge, appropriate ICT tools need to integrate and process large volumes of data generated daily. Into this end, the Big Data concept has become a hype in both industry and academia over the last years, and can be seen as broad terminology for extremely large and complex data sets, which cannot be adequately handled by traditional data processing tools and mechanisms (Zhu, Yu, Wang, Ning, & Tang, 2019).

The exponential growth of data, generated from multiple sources, results in the formation of what is currently known as Big Data. Data sources are around us everywhere, smart phones, computers, environmental sensors, cameras, GPS (Geographical Positioning Systems), and even people. Therefore, there is a huge potential in adopting a Big Data approach to address mobility-related problems directly from the source as well as analytics for deeper insights through data analytics, data intelligence and deep learning. The benefits of applying Big Data approaches to transportation-related challenges can be described as follows:

- Generation of new traffic patterns by analyzing real time data
- Reduction of congestion levels, by predicting traffic conditions and adjusting congestion charging strategies
- Reduce supply chain waste by integrating real-time data to deliveries and thus, optimizing shipping routes.
- Enable data streaming to process and communicate traffic information collected through sensors, smart traffic lights and on-vehicle devices to drivers via smartphones or other communication devices.

In summary, we advocate that the adoption of Big Data approaches to transportation-related challenges, can be seen as important mechanism to provide real-time insights for specific entities to take actions to alleviate or resolve a traffic problem.

The objective of the work described here is to contribute to the reduction of congestion levels, by proposing a Big Data approach, with which toll prices (currently static in existing highway infrastructures) may be dynamically calculated, near real-time, to cope with congestions and traffic issues, while ensuring the maximum revenue and the quality of the traffic service in the road infrastructure, i.e. the Level of Service (LoS). This is called congestion charging (Button, 1993). Congestion charging is used to promote an efficient usage of existing road networks while, at the same time, reducing congestions levels, leading to lower commuting times and improving the quality of life of commuters. The topic of congestion charging, namely dynamic congestion charging, will be further explained in Section 2.

An efficient dynamic congestion charging system should be supported by a Big Data technology stack that is able to process ever-expanding quantities of data, collected from sensors scattered along a big geographic area, in near real-time, in order to enable high dynamics in toll pricing decision making so as to cope with high traffic variability scenarios, such as traffic congestion or accidents.

1.1. Motivation: The case of shadow-tolls in Portugal

Portugal has witnessed great investments on highway infrastructures over the last 30 years, with an inland global transport infrastructure investment of 1.2% of Portugal's GDP between 1992 and 2010 (OECD, 2013). The OECD average is approximately 0.8%, excluding Japan. In 2015, the highway network was six times longer than in 1990, as the length of highway per capita grew above the EU average (Figure 1 b)) (OECD, 2015).



Figure 1. a) Evolution of highway kilometers per country area, from 1990 to 2014 in Portugal; b) Number of highway kilometers per 1 million inhabitants in Portugal, compared with other countries

These huge investments were mainly public investments, partly from European Regional Development and Cohesion Funds. Nevertheless, in the 90's, the Portuguese government decided to call for the participation of private companies in this process (Instituto de Mobilidade e dos Transportes, 2014) in an attempt to accelerate the development plan and transfer the initial investment and associated risks of the highway operation to the private sector. Hence, a new concept of Public-Private Partnership (PPP) scheme was introduced to compensate the private sector for investing in less appealing road investments, while promoting free-access to the highway network in social and economic deprived areas of the country. This model is known as “shadow-tolls” payment, formally named by the Portuguese government as SCUT highways, the Portuguese acronym for “no cost for users” (Santos & Santos, 2012), (Unidade Técnica de Acompanhamento de Projectos, 2017). Simultaneously, Infraestruturas de Portugal (IP), a public limited company, was formed through the extinction of a former national road agency to operate these roads. IP became the main investor in the road network through a concession contract signed with the government, and the main responsible for the SCUT highways' PPP (Instituto de Mobilidade e dos Transportes, 2014).

Although the application of the SCUT scheme could be justified under certain circumstances, the program eventually gave rise to financial constraints. Thus, in 2007, the Portuguese government decided to re-evaluate the SCUT program, eventually canceling this scheme and allowing concessionaires to charge road users (Santos & Santos, 2012). The SCUT shadow-toll collection was introduced in October 2010. Each toll has a fixed value set on every concession contract, and annually updated, in the first month of each calendar year, in accordance with a calculation mechanism. A different toll value is set for each type of vehicle, based on the distance to charge and including a value added tax (VAT). The toll is charged through a Multi Lane Free Flow (MLFF) system, an innovative technology and suitable for installation on already existing highways that were originally designed to be toll-free (Santos & Santos, 2012).

However, the introduction of tolls in SCUT highways has had a significant negative impact on travel demand (and consequently on the mobility and accessibility of certain regions), with travel cost increases of almost 50 percent in certain highway sections, with respective consequences in terms of greater emissions, travel times and overall lack of satisfaction of the road users (Bandeira, Coelho, Pimentel, & Khattak, 2012). Another constraint of this scenario is that the system does not cover the full extent of the network, as not all highways sections have MLFF toll gantries installed due to investment restrictions associated with the introduction of a toll collection model during the most significant extension of the network. This promotes an erratic and inefficient usage of highways by users that ultimately circulate only on toll “free” sections, fleeing to the alternative national roads on the rest of their routes (Wise, 2013). As a result, highways are underused and the alternative national roads are highly congested, while IP is left dealing with two major questions:

- The national roads were not planned to have such a high traffic demand, especially that of heavy vehicles, and therefore deteriorate faster than expected, resulting in higher maintenance costs (Wise, 2013).

- The massive investment that was made on the highway network is not yielding the expected return as the traffic is lower than initially planned resulting in lower revenues (Unidade Técnica de Acompanhamento de Projectos, 2017).

Hence, IP has been trying continuously to devise strategies to recover traffic, optimize road network usage and allow greater efficiency of the model to be reached. Two types of discount were applied on highways with IP's tolls revenues. In 2012, a universal 15% discount was issued on old SCUT's, in order to continue to ensure that the impact associated with the introduction of toll collection would be largely mitigated and at the same time guarantee there's no discrimination between highway users (Governo de Portugal, 2012). Also in 2012, lined up with the principles set out in "Eurovignette" framework (Directive 2011/76/CE) (European Union, 2011), a modulation scheme for Heavy Goods Vehicles (HGV) was introduced, in which freight transport may benefit from a 10% discount on daytime period and 25% at night and weekends, encouraging traffic shift to less demanding hours. On August 2016, an additional 15% universal discount was applied to certain concessions, in particular those which highways serve regions with poorer economic indicators. Additionally, the HGV scheme was reinforced by increasing the discount to 15 and 30%, in each period, and extending night period (Governo de Portugal, 2016).

Regardless, the HGV scheme has not had much acceptance, which convey the lack of customized solutions, better serving freight sector needs. In addition to this, the second universal discount measure turned out to have a negative impact on toll revenues so far (Instituto de Mobilidade e dos Transportes, 2014), (Wise, 2013). Therefore, a more commercial approach, providing more flexibility on pricing, is apparently needed. Focusing solely on HGV, the introduction of tolls on former SCUT's had several consequences, both economical and operational. On the economical side, despite the cost of using national roads being high, not only in terms of fuel costs but also regarding vehicle maintenance costs, not to mention increased travel time and less comfort and security, freight operators prefer not to use tolled highways and go through national roads instead, since the toll cost is higher than the whole cost of choosing national roads. On the operational side, and since some stretches of highway are toll-free, the selected routes normally use these stretches whenever possible, and deviate to national road alternatives whenever a stretch is tolled (Audretsch, Dohse, & dos Santos, 2017). This creates a zigzag pattern for the freight routes, which can also have more costs associated.

Hence a new approach that could shift traffic from alternative national roads to shadow-toll highways, namely HGV traffic is needed since heavy vehicles are responsible for a big part of the maintenance costs due to road pavement deterioration, which reduce safety conditions on such roads. A behavioral change in travelling choices, is expected to benefit both the road operators, not only by decreasing maintenance costs, but also through an increase in toll revenues, and freight operators, which would shift their operational routes to tolled highways, benefitting from proper discounts/price reductions while saving in fuel and vehicle maintenance costs. To achieve the aforementioned aim, it is necessary to create effective mechanisms and incentives which enable freight operators to change their fleet management operations towards a better usage of highway infrastructure, considering the toll price fluctuations. These conditions encompass the creation of competitive prices depending both on the journey length and in the time spent.

1.2. Main contributions

The presented work has been developed under the European Commission-funded H2020 OPTIMUM research project (OPTIMUM consortium, 2015) and proposes a dynamic toll charging system for shadow-toll highways in Portugal, supported by Big Data technologies, so as to induce changes in HGV behaviour by diverting heavy vehicle traffic from urban and national roads to underused tolled highways. This is accomplished by attracting or discouraging the use of specific highways through toll prices' variability, according to the quality of service prediction on those highways and adjacent alternative roads.

The system is fed by traffic flow conditions of both tolled highways and their national road alternatives, combining historical and real-time data collected from traffic sensors scattered throughout highways and alternative national roads, in order to calculate the toll pricing of highways in advance, depending on traffic

congestion conditions on both road types. The design and development of the dynamic toll pricing model takes into account the traffic flow now-casts (including traffic events, maintenance, accidents and weather related situations) and traffic flow forecasts, resulting in more accurate predictions for highways and national roads. Since traffic data quantity and quality are crucial to the prediction of road networks' statuses, real-time and predictive Big Data analytics methods are used.

Hence, the dynamic toll charging system needs to be supported by the latest Big Data technologies in order to efficiently collect and process big amounts of traffic data, and swiftly perform traffic now-casts and forecasts to feed the dynamic charging model. Therefore, the main contributions of the work presented here can be highlighted as follows:

- Development of a Big Data infrastructure capable of collecting and processing large volumes of traffic data in real-time, and swiftly perform forecasting analytics to produce traffic predictions;
- A mathematical model for dynamic toll price calculation, which takes into account both the traffic in the tolled highway and in its toll-free, alternative roads;
- Integration of real-time toll pricing results of the dynamic toll charging system with logistics operator fleet management systems through a novel User Interface for dynamic toll information;
- Test and validation of the system in a real-world scenario, targeting HGV.

The main contributions beyond state of the art of the presented work reside on both the application of Big Data technologies on toll charging systems, and on the way the dynamic charging model is conceptualized. For the best of our knowledge, and besides previous works from the authors, the application of Big Data technologies on dynamic toll charging systems is inexistent in the literature, as will be evident in Chapter 2. Furthermore, most, if not all, real-world dynamic toll charging systems calculate the dynamic price of the toll depending solely on traffic, environmental or weather conditions in the specific highway in which the toll is being charged (e.g. High Occupancy Toll lanes), while in the literature several dynamic charging models for linked networks are presented theoretically, most are not tested, validated and updated in a real world scenario, and therefore do not take into account the knowledge of real road infrastructure operators', as will be shown in Section 2.

The OPTIMUM dynamic charging system contributes to the development of the existing state of the art with a novel paradigm in which the toll price calculation depends not only on the traffic in the specific highway in which the toll is being charged, but also on the traffic conditions of national roads that are common alternatives to the tolled sections of the specific highways. In addition, it presents a real-world scenario, which is supported and validated by a real public road infrastructure operator, IP.

The OPTIMUM dynamic toll charging system enables freight operators to plan their routes according to the dynamic variation of tolls. This is important both from the freight operators' perspective, as it will enable them to perform their activity with higher standards of efficiency, safety and speed. From the road infrastructure operator's point of view, it would minimize the maintenance costs of national roads while taking into account the current vehicle load, namely HGV load, allowing for an appropriate use of highway and national road infrastructures, while increasing toll revenues. Specifically, the OPTIMUM dynamic toll charging system should:

- Shift freight traffic from urban and national roads to tolled highways (15%);
- Increase toll revenues for the road operator (10%), and;
- Improve the efficiency of operations for freight operators (10%).

This is accomplished by comparing the Level of Service in tolled sections of highways and alternative national roads, in terms of traffic flows and congestion, through the application of a mathematical model and apply a discount to the fixed value of the toll. The dynamic charging system was tested and validated by logistics company Luis Simões (LS), one of the biggest freight operators in the Iberia Peninsula, and by IP, the Portuguese road infrastructure operator. The mathematical equations for the toll charging model were developed, refined and validated during two real-world pilot iterations.

The remainder of this paper is organized as follows: section 2 reviews the literature regarding Congestion Pricing, with special attention to Dynamic Toll Charging, and also the application of Big Data technologies on toll charging systems. Section 3 presents the overall OPTIMUM dynamic toll charging system architecture and Section 4 provides a brief overview of the OPTIMUM dynamic toll charging model, developed for Portuguese shadow-toll highways. Section 5 thoroughly describes the Big Data tools and services that feed the dynamic toll charging model, while Section 6 presents and discusses the validation results of the real-world pilot test iterations performed in the scope of the OPTIMUM project.

2. Literature Review

This section presents the concept of Dynamic Congestion Charging and highlights the main academic works and real-world implementations about the topic. Furthermore, a literature review on Big Data-supported toll systems is presented. In both cases, the major progresses beyond the current state of the art are described.

2.1. Dynamic Congestion Charging

2.1.1. Congestion Charging

Congestion charging can be defined as the process of "charging drivers within peak times to persuade them either to change their travel times or to use alternative routes that are not overloaded at peak hours" (Yildirim, 2001). The theory behind road pricing entails that to achieve social optimum, a toll is required to be levied in proportion to the difference between social marginal costs (which includes external costs that users impose on an overloaded road) and private average costs (travel delays, fuel, maintenance, etc.) (Shirazi, Aashtiani, & Quadrifoglio, 2017). By charging the users, transportation planners try to fulfil one or more of the following objectives (Yildirim, 2001):

- *Cost recovery for planning, construction, operation and maintenance.* Widely used and common objective to recover the building cost of new infrastructures, such as highways, tunnels or bridges. In the congestion pricing context, users can be charged by road managers for the extent of the congestion that they are introducing on the transportation network.
- *Retrieval of negative externalities.* In this case, besides the direct costs of the trips like fuel and travel time, the users must also pay for other impacts such as the additional travel delay, air pollution and accidents that may impose to others (Fathollahi-Fard & Hajiaghahi-Keshteli, 2018), (Sahebjamnia, Fathollahi-Fard, & Hajiaghahi-Keshteli, 2018). These costs are commonly defined as negative externalities.
- *Demand management.* In order to obtain an optimum utilization of the infrastructure, that is, try to keep a balance in the traffic volume among all points of the road network, some segments can be tolled to shift demand and influence users' routes choices, shifting demand to underutilized areas.

In order to fulfil the two last objectives, congestion charging is seen as one of the most effective tools to mitigate traffic congestion (Liu, Wang, & Meng, 2014). Congestion charging comes in two different flavors: static congestion charging and dynamic congestion charging (Morgul, 2010), (Xu, 2009):

- *Static Congestion Charging.* A toll system in which rates depend on the time of day. The system is static because the toll rate schedule does not reflect the actual traffic conditions, and often remain static for long periods of time. The initial batch of surveys on congestion pricing have mainly considered this type of "static" charging, focusing mainly in the optimization of toll locations, toll prices and the optimal places to apply tolls in a large road network (e.g. (Walters, 1961), (Vickrey, 1969), (Beckmann, 1965)).
- *Dynamic Congestion Charging.* A toll system in which the price of the toll varies in near real-time, according to the traffic conditions. Various traffic parameters can be taken into account to calculate toll rates, including road occupancy, speed, or delays in traffic. These parameters can be measured in real time and the toll fee updated in short time intervals. With the help of variable message boards, smart phone notifications or even web interfaces, drivers are informed about the current charge rates and can use the tolled road/lane to save time or choose a free, more congested alternative. Research on dynamic congestion pricing has increased markedly in recent years (Yu B. , Zhang, Guan, Peng, & Yao, 2017). Dynamic tolls can take many forms according to the time of day, road type, vehicle characteristics and even the current traffic conditions (e.g. (Li, Saigal, & Zhou, 2012), (Cheng & Ishak, Maximizing Toll Revenue and level of service on managed lanes with a dynamic feedback-control toll pricing strategy, 2016), (Hourdos, Janson, Levinson, & Parikh, 2015), (Gutman, 2016)).

2.1.2. Real-world implementations of congestion charging

Throughout the years, there have been many attempts to implement congestion charging on road and highway infrastructures. Singapore was the first city to adopt such a scheme, back in 1975, in order to reduce traffic and improve air pollution (Seik, 1997). The scheme was called Singapore's Area Licensing Scheme (ALS), and it charges fees depending on time of day, to enter the restricted zone in the central area of Singapore. Norway had success with road pricing in three urban centers: Oslo, Bergen and Trondheim. For instance, in Trondheim where toll booths operate with an electronic card system, the rush hour traffic has decreased by 10% while non-peak period traffic has increased by 9% (Yildirim, 2001).

In the State of Oregon, a pricing scheme based on per-mile charges was tested, which will replace fuel taxes in the near future. The used congestion-pricing component produced higher toll rates during periods of high traffic volume on specific road segments. The Puget Sound Regional Council studied the travel behavior of a family when applied a similar charging system in the Seattle metropolitan area, from 2005 to 2006. Charges were mainly calculated considering the type of facility and its level of congestion (Federal Highway Administration, 2017).

Cordon pricing consists on charging drivers as part of a demand management strategy to soften traffic congestion within an area (e.g. city center). There are several cordon deployments in Europe and Asia (e.g. London, Rome and Stockholm) (Morgul, 2010). The London congestion charge was introduced in 2003, and since then has been closely monitored. Its fifth annual report states that the gross annual income for the original scheme was £200 million and the total costs were £88 million, originating a net income of £112 million and a benefit–cost ratio of 2.27. Stockholm's congestion charge was born as a 7-month trial and became permanent in 2007, after a successful referendum on the matter (de Palma & Lindsey, 2011).

To achieve a more efficient use of highways, many states in the United States (U.S.) have either built or reserved special lanes for carpools - vehicles with more than one passenger. These specific lanes are known as high occupancy vehicle (HOV) lanes (Li J. , 2001) and are intended for high occupancy vehicles such as public transit buses and emergency vehicles, which are allowed to use the lanes free of charge or at reduced rates. While some HOV lanes have been proven effective, others are found to be underutilized, because carpools account for only a small proportion of total vehicle travel and not all car-poolers use HOV lanes.

Since 1995, some U.S. states have successfully implemented High Occupancy Toll (HOT) lanes to manage high congested road sections. HOT lanes are alternately described as managed and express lanes. HOT lanes offer an alternative to try to mitigate the problem of underutilized HOV lanes (Murray, 2012), (Gutman, 2016). Hence, several works describe the conversion of HOV lanes into HOT lanes (Li J. , 2001), (Naga, 2007), (Lou, Yin, & Laval, 2011). In general, HOT lanes allow single occupancy vehicles (SOVs) to pay a toll for using the HOV lanes. This is the case of Colorado, Georgia, Florida, California, Texas, Minnesota, Texas, Virginia, Utah and Washington. Each of these nine U.S. states have developed tailored made solutions to fit their traffic movements and freeway characteristics, with the objective of improving the travel experience and offer alternatives to commuters. A review of all HOT lanes in the U.S. can be found in (Morgul, 2010), (Los Angeles County Metropolitan Transportation Authority, 2017).

Finally, the Fast Lane (FL) in Israel is a HOT lane parallel to Highway 1 (HW1) with 11.4 km long, westwards from Ben Gurion Airport (BG) to the Kibbutz Galuyot interchange at the entrance to Tel Aviv. The FL has a toll fee determined by a Level-of-Service algorithm, which depends on the average speed on the tolled sections. The average speed must be at least 70 km/h, which means that drivers have to cross the FL in 10 minutes. If this objective is not achieved, then the driver does not pay the toll fee. It works as a subscription service, with users being identified through a number plate by an identification system. Subscribers pay the toll automatically via bank transfer; non-subscribers can pay at a nearby Park-and-Ride plaza (Gutman, 2016).

2.1.3. Advantages of Dynamic Toll Charging

Dynamic toll charging has been introduced through the use of HOT lanes. The use of such toll charging systems reduces traffic congestion but also allows for price discrimination to arise as users are charged

different prices dependent upon the type of vehicle used. This is given further emphasis if the firms that implement dynamic toll charging operate in monopoly market structures, a market structure where there is a single prevalent firm supplying the market (Deck & Kimbrough, 2014).

Furthermore, highways are often denoted as a common pool resource. This refers to a resource whose size and specific characteristics make it harder to exclude potential beneficiaries from attaining benefits from its use. Therefore, such resources often face issues of congestion, which means that the value of the negative externalities present is large. The presence of negative externalities arises as a direct consequence of traffic congestion (Winaisathaporn, 2013). As a result, there are external costs to the economy presented in the form of increases in pollution and additional emissions caused by greater traffic congestion (Fathollahi-Fard & Hajiaghahi-Keshteli, 2018). However, the implementation of dynamic toll charging can be looked upon as a viable solution for the correction of the negative externalities through a decrease in the level of pollution and emissions brought about by a reduction of traffic congestion, which makes society better off.

One of the predominant benefits regarding the implementation of dynamic toll charging is that it causes drivers to internalize the externalities, which arise as a result of traffic congestion. Therefore, only individuals who value a specific road due to factors such as time constraints, will choose to do so as opposed to using alternative routes. The internalization of such externalities as a by-product of the use of dynamic toll charging can have several impacts on the economy. For instance, (Morgul, 2010) found that the implementation of dynamic toll charging in South Florida increased the use of public buses within the first six months of the tolls being implemented whilst generating over \$2.8 million in revenue, which accounted for 89% of its total projected value.

From an economic standpoint, the implementation of dynamic tolls can also be looked upon as an efficient way to decrease traffic congestion. In addition, the implementation of dynamic tolls yields the benefit of decreasing traffic congestion costs, which have accounted for an estimate of 1.5% of GDP in the UK and 1.3% and 0.9% for France and Germany. In 2014, traffic congestion caused Americans to travel an additional 6.9 billion hours and purchase an additional 3.1 billion gallons of fuel resulting in a total congestion cost of \$160 billion on the US economy, approximately 0.9 per cent of US GDP. Trucks accounted for 17 per cent of that cost, approximately \$28 billion, whilst accounting solely for 7 per cent of traffic (The Texas A&M Transportation Institute, 2015). This is a value that could have been immensely reduced with the implementation of dynamic tolls making society better off as a whole.

2.1.4. State of the art on Dynamic Toll Charging

HOT lanes are the existing systems closest to pure dynamic toll charging. This means that, although some of them work as time-of-day tolls, the price of a HOT lane can change in real-time depending on the actual traffic conditions of the highway in which the system is installed. Purely dynamic, responsive toll charging systems are only theoretical (Shirazi, Aashtiani, & Quadrifoglio, 2017), (Zhang, Zhang, & Chen, 2019), and are not deployed, tested and validated in real-world applications or reflect the real purposes and knowledge of a real road infrastructure operator (Shirazi, Aashtiani, & Quadrifoglio, 2017), (Zhang, Zhang, & Chen, 2019), (Bracher & Bogenberger, 2017).

On the other hand, the *modus operandi* of congestion pricing schemes is often to raise prices depending on the traffic congestion on a certain road or area (Li, Saigal, & Zhou, 2012), as in the case of HOT lanes (e.g. (Cheng & Ishak, 2016), (Hourdos, Janson, Levinson, & Parikh, 2015), (Gutman, 2016), (Leonhardt, Saches, & Busch, 2012), (Jang, Chung, & Yeo, 2014)), while it could also lower prices depending on other factors, such as the traffic in alternative, toll-free roads. Even so, and to the best of our knowledge, there are few works in the literature, or none at all, in which the congestion charging system on a particular road/highway depends on traffic conditions in other, often alternative, roads in order to lower/raise the price for that specific road/highway.

The proposed dynamic toll charging strategy and model offer a novel approach to purely responsive dynamic toll charging, supported by real-world knowledge from a public road infrastructure operator, in this case IP, the Portuguese public road and highway operator, which enables the dynamic calculation of toll prices,

depending on both the traffic conditions of tolled highways and their toll-free, national road alternatives. Opposite to HOT lanes, the main idea is that if the traffic conditions deteriorate in the alternative roads, the toll price should be lowered so as to deviate traffic from national alternatives to tolled highways. If, in contrast, the traffic conditions in highways fall below the required level of service threshold, the toll price should go up, in order to encourage drivers to take the toll-free alternative.

2.2. Big Data-supported toll systems

Recent years have witnessed a dramatic increase in our ability to collect data from various sensors, devices, in different formats, from independent or connected applications. According to a report from the International Data Corporation (IDC), in 2011, the overall data volume in the world was 1.8 Zettabytes ($\approx 1021\text{B}$), increasing by nearly nine times within five years (Addo-Tenkorang & Helo, 2016). This data “tsunami” (Zhong, Newman, Huang, & Lan, 2016) has outpaced our capability to process, analyze, store and understand these datasets. The explosion of mobile networks, cloud computing and new technologies has given rise to incomprehensibly large amounts of information, often described as “Big Data”. Big Data is a broad terminology for extremely large and complex data sets, which cannot be adequately handled by traditional data processing tools and mechanisms (Fahmideh & Beydoun, 2018).

Hence, Big Data can be seen both as highly heterogeneous, voluminous data sets that may possess near-real time, high speed granularities, and to the newly available technologies that are made specifically to handle such type of data, although there are other means to solve the problem of data complexness, such as the use of heuristics and meta-heuristics (Haijaghaei-Keshteli & Fathollahi-Fard, 2018), (Sahebjamnia, Fathollahi-Fard, & Hajiaghaei-Keshteli, 2018). Such technologies attracted great attention from industry and academia, with some of the bigger data-related companies, such as SAP, IBM, Oracle or Microsoft, spending over \$15 billion on Big Data analytics and processing technologies, with an expected market value over \$118 billion in 2022 and growth prospects of 10% to 26% per year, from 2014 to 2022 (Zhong, Newman, Huang, & Lan, 2016).

Furthermore, recent literature shows that several industry fields are using, or interested in the implementation of Big Data-handling systems to handle their ever growing datasets. From manufacturing (Addo-Tenkorang & Helo, 2016), (Zhong, Newman, Huang, & Lan, 2016), (Fahmideh & Beydoun, 2018) to health (Lee & Yoon, 2017), and encompassing sales (Li, Ch'ng, Chong, & Bao, 2016), (Liu J.-W. , 2018), maintenance (Dinis, Barbosa-Póvoa, & Teixeira, 2018) and insurance (Fang, Jiang, & Song, 2016), Big Data is now widely adopted and is being applied to several existing data-related problems, with significant financial and commercial implications (Yu & Zhang, 2017). Big Data is also becoming a prominent research focus in the field of ITS (Zeyu, Shuiping, Mingduan, Yongqiang, & Yi, 2017), and several works have been done and are present in the literature (Zhu, Yu, Wang, Ning , & Tang, 2019).

Even so, within the context of toll systems, and apart from the work done in the OPTIMUM project (e.g. (Guerreiro, Figueiras, Silva, Costa, & Jardim-Gonçalves, 2016), (Figueiras, et al., 2016) , (Figueiras, Guerreiro, Silva, Costa, & Jardim-Gonçalves, 2018), (Petalas, Ammari, Georgakis, & Nwagboso, 2017), (Ramos, et al., 2018)), there is a limited number of works worth mentioning. In (Kumar & Machado, 2018), the authors propose a tollgate system enhanced by Internet of Things and Big Data technologies, in order to minimize traffic density, reduce stop time at toll gates and detect frauds. The system uses sensors to electronically debit the accounts of registered car owners, and alert law enforcers for those that are not. (McQueen, 2017) proposes toll charging and electronic toll payment as one of the future trends in Smart City-related services.

As already highlighted, the proposed system progresses beyond the current state of the art by applying Big Data technologies to a dynamic toll charging system. This application results in a Big Data platform for dynamic toll charging, capable of swiftly and efficiently collecting, harmonizing and cleaning big volumes of data, and then rapidly and accurately perform traffic forecasting with such data volumes, in order to feed the dynamic charging model.

3. OPTIMUM Dynamic Toll Charging System Architecture

The H2020 OPTIMUM research project (OPTIMUM consortium, 2015) is set on a ubiquitous connectivity environment within the transportation system and everything that surrounds it, continuously providing data on the current state of the transportation system, and on any emerging situation or issues that can arise. Data is gathered from a panoply of sources, such as road sensors, in-vehicle sensors including positioning information (e.g. GNSS data), crowd data sourcing through social networks, occupancy of public transportation and availability of transportation modalities like shared cars and shared bicycles.

Information management and gathering is supported through the development of a scalable, highly distributed architecture for managing and processing Big Data from multiple sources, while enabling continuous monitoring and control of the transportation system and providing proactive, new data-driven mobility services in a (semi-) automatic and personalized manner. The OPTIMUM platform's architecture encompasses several Big Data-based tools, backed up with different technologies, which not only provide the necessary services to the project's pilot cases, but can also be reused in new mobility scenarios. The conceptual architecture is supported by the Observe-Orient-Decide-Act (OODA) loop, as proposed by John Boyd (Osinga, 2005).

Research developed under the Portuguese pilot of OPTIMUM aims at contributing with solutions for both freight and highway infrastructure operators' tolled highway-related issues, enabling profitability and best route choices through the use of novel ITS and Big Data services. The idea is to capitalize on the intelligence that can be derived from the abundance of transportation-related data in order to create a responsive dynamic toll charging system to optimize the usage of highways and national roads which offer a toll-free alternative to these highways. The specific architecture for the OPTIMUM Portuguese pilot case is shown in Figure 2.

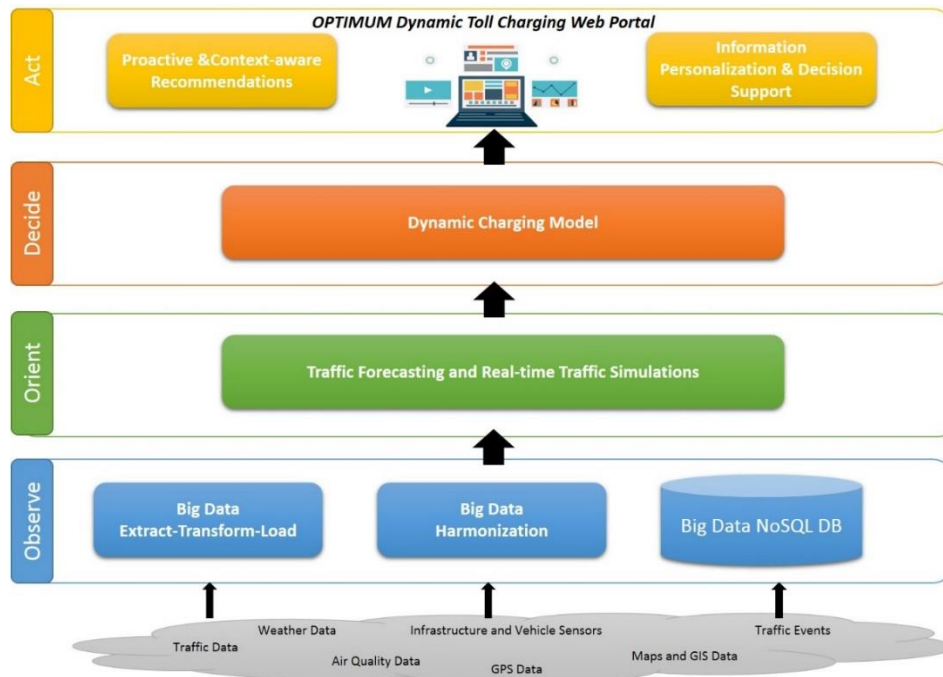


Figure 2. OPTIMUM Project's conceptual architecture – Portuguese Pilot

In the Observe stage (Section 5.1), OPTIMUM provides the required Big Data infrastructure to handle and process large volumes of ITS-related data, characterized by high refresh rates and high data heterogeneity. The Big Data infrastructure has specific adapters to support data collection from several data sources. In the case of the Portuguese pilot, collected ITS-related data spans from traffic data from road sensors, collected in near-real time, to weather data, among others.

Regarding the Orient stage, services for situation assessment enable the generation of real-time, data-driven predictions, as well as the discovery of anomalous conditions, based on events from Complex Event Processing tools of forecasting and now-casting services. In the context of the OPTIMUM's Portuguese pilot, the OPTIMUM traffic forecasting engine (Section 5.2) allows fusion of historical and real-time information, retrieved from multiple sources, in order to provide traffic forecasts for different time granularities (e.g. every five minutes, hourly, etc.), both in the tolled sections of highways and their alternative, toll-free national roads.

In the Decide phase, OPTIMUM explores the feasibility of novel charging and crediting schemes, where users are "debited" based on the use of the transport network, or the environmental footprint of their trips. These charging and crediting schemes enable the implementation of eco-friendly trip options, in order to solve issues such as pollution hot-spots along the transportation networks. Within the scope of its Portuguese pilot, the OPTIMUM project provides a dynamic toll charging strategy dedicated to mitigate situations such as the degradation of the road infrastructure and the decreasing level of service in highways, due to the impact caused by freight vehicles (Section 4). The results are variable toll prices for freight vehicles, based on travel distance, vehicle classification (e.g. size, emission outputs), maintenance costs and the actual predicted traffic situation.

Finally, the Act phase comprehends a repository of personalized persuasive strategies, which enable the delivery of personalized information to end users, both in terms of form and content. Such strategies are supported by psychological principles that rule persuasion tactics, interface design strategies, and other variables with the potential to influence behavior in general. The objective of the deployment of such strategies is to nudge user mentalities towards system-wise optimal behaviors or actions that improve or enable sustainable mobility practices. For the use case on dynamic toll charging, the Act phase consists in decision-making support for freight operators, through personalized, context-aware recommendations in the form of a toll pricing presentation interface.

4. Overview of the Dynamic Charging Model

In order to construct a dynamic toll charging strategy that could meet the revenue needs of the road infrastructure operator whilst enabling freight operators to use tolled highways affordably, several steps were followed. First, a traffic assignment and modelling process was conducted in terms of supply, characterizing road network systems within the pilot's area. The traffic assignment task estimates the traffic volumes, travel costs and travel times for each link and demand between Origin-Destination (OD) pairs, taking as input the description of the transportation system and an OD matrix of trip demands. To build the OD matrix, data from four trucks was provided by the freight operator, LS, spanning 224 days, and about 3.600 trips during the entire period of 7 and a half months were obtained. The trucks were carefully chosen from the total freight fleet, enabling a reliable characterization of the routes in the pilot network area. The pilot area was comprised by several highways in the North of Portugal, which provided an accurate representation of the Portuguese shadow-toll highways and their free-toll alternatives.

To estimate traffic volumes and travel times, and in conjunction with information extracted from the Luis Simões' OD matrix, surveys to freight operators and HGV drivers were conducted, in the form of questionnaires, to gather additional information about travel times, behaviors, costs and other variables that were important to consider from the freight operators' perspective. Questionnaires were also handed out to road concessionaires and to the road infrastructure operator, IP, in order to assess important factors for the charging strategy, such as road infrastructure maintenance costs, both on national roads and highways, toll revenue streams, and possible revenue deviations due to incentives and toll pricing changes.

Next, a route choice model, based on a binary, discrete-choice logit model was created. Route choice is extremely important for dynamic congestion pricing, since every commuter makes his/her daily travel choices and can alter or switch his/her habitual route to avoid congestion or to avoid a toll. This also happens in the case of freight drivers. Therefore, route choice must be considered in every congestion pricing model. In route choice studies and models, the main issues are normally associated with how to simulate accurately and realistically drivers' choices (Xu, 2009). The author of (Xu, 2009) further elaborates on route choice models, describing two of the most used methods for route choice models: fuzzy logic and discrete choice models, also presenting some research works on these two methods.

In this model, the distribution model was the logit model (Liu, Wang, & Meng, 2014): the distribution rule is based on the difference between the generalized costs (impedances) of each route. The traffic model calculates, for each alternative route, the generalized cost of the trip, which is used as "utility" to distribute demand between different routes. The main assumption is that an individual will choose the alternative with the highest utility (Ben-Akiva & Lerman, 1988) and social impacts important to him/her (Fathollahi-Fard, Hajiaghahi-Keshteli, & Mirjalili, 2018). Equation 1 shows the explicit expression of the probability of a freight forwarder, n , choosing route, j :

$$P_{nj} = \frac{1}{1 + e^{-\mu V_{nj}}} = \frac{1}{1 + U_{nj}} \quad (1)$$

Where,

P_{nj}	Probability of route j , chosen by freight operator n
V_{nj}	The generalized cost (or impedance) of route j , chosen by freight operator n
μ	Logit model scale parameter, normalized to 1
U_{nj}	Utility of route j , chosen by freight operator n

A sensitivity analysis was carried out in order to achieve the best scale parameter for the logit model, and a value of $\mu = 1$ was chosen since it provided the most realistic distribution results. The utilities of the logit route choice model are the given by the two (binary) equations below:

$$U_1 = \beta_0 + \beta_{traveltime1} * TravelTime_1 + \beta_{TravelCost1} * Cost_1 \quad (2)$$

$$U_2 = \beta_{traveltime2} * TravelTime_2 + \beta_{TravelCost2} * Cost_2 + \beta_i * Incentive_i + \beta_j * Variable_j + \varepsilon_{toll} \quad (3)$$

Where,

U_1, U_2	Utility for national road and tolled highway, respectively
$TravelTime_1$	Travel time of alternative 1
$TravelTime_2$	Travel time of alternative 2
$Cost_1$	Travel cost of alternative 1 (Fuel cost for national road)
$Cost_2$	Travel cost of alternative 2 (Fuel cost and toll cost for tolled highway)
β_0	Alternative road specific constant
$\beta_{traveltime*}$	Travel time coefficient for the respective utility
$\beta_{travelCost*}$	Travel cost coefficient for the respective utility
β_j	Variable coefficient
ε_{toll}	Gaussian, zero-mean error term, with a standard deviation

The estimation of all the above coefficients was carried out with the help of the freight operators and drivers' questionnaire results, since they are based on travel-related characteristics, such as travel time and cost, as well as the possibility of existence of specific incentives. Such estimation enabled the calculation of the Value of Time (VoT) for LS freight transport operations, both in terms of the operator itself and its drivers. Table 1 presents the result of the coefficient and VoT estimation. The positive sign of the value of the estimated alternative specific constant shows that, *ceteris paribus*, there is inertia towards the tolled highway alternative. The coefficients of travel time and travel cost have the expected signs, and the latter is statistically significant as expected.

Model	Only Operators		Only Drivers	
Variable	Coefficient	t-stat	Coefficient	t-stat
Constant <i>National</i>	1.62	1.24	-1	-0.93
$\beta_{TimeNational}$	-0.141	-3.56	-0.0267	-2.10
$\beta_{TimeToll} (\beta_{tt})$	-0.168	-2.46	-0.0551	-2.33
$\beta_{Cost} (\beta_{tc})$	-0.204	-3.87	-0.257	-4.83
$\beta_{CostToll}^*$	-	-	-0.156	-4.42
$\beta_{NumberOfTrucks}^*$	0.0104	2.10	-	-
$\beta_{RestingArea}^*$			0.212	1.99
$\beta_{DedicatedLane}^*$			0.0155	2.58
$\beta_{PenaltyforDelay}^{**}$	-0.883	-2.71		
ε_{toll}	1.17	3.82	-0.969	-4.88
*Specific to Toll				
** Specific to National				
rho-square	0.540		0.552	
Observations	265		417	
VoT - National	41.5		10.3	
VoT - Toll	49.4		21.19	

Table 1. Logit model utilities' coefficient estimation results

VoT, or the users' willingness to pay for a unit change in travel time is one of the important factors for determining users' routes and time departure choices (Morgul, 2010). Several VoT studies were conducted for passenger trips in different regions of the world. In the case of commercial VoT models, there is a more

limited amount of research available (Morgul, 2010). The model estimation results enabled the calculations of VOT in relation to trucks operated by freight operators. Equation 4 was used to calculate the VOT for freight operators.

$$VoT_{fo} = \frac{\beta_{tt}}{\beta_{tc}} = \frac{-0.168(\text{minutes})}{-2.04(\frac{\text{€}}{\text{minute}})} * 60 = 49.4 \text{ €/hour} \quad (4)$$

The calculated VoT is similar to the ones calculated by local studies in the study area, and is also consistent with the findings of the UNITE (UNification of accounts and marginal costs for Transport Efficiency) project (European Commission, 2003), when updated for the inflation values after 2015.

The next step was to calculate travel times in each section, both on highways and national roads, depending on the delay caused by the number of vehicles present in the specific road section. Since, the overall capacity of both highways and national roads is not often surpassed, the BPR volume-delay function (Kucharski & Drabicki, 2017) was used to calculate travel times in both road types (Equation 5).

$$t_{cur} = t_0 * \left(1 + a * \left(\frac{q}{q_{max} * c} \right)^b \right) \quad (5)$$

Where,

t_{cur}	Current travel time for flow q
t_0	Free flow travel time
q_{max}	The capacity of the road network
a, b, c	Parameters set by the model designer/expert

After several tests, parameters a , b and c were adjusted to mimic the average travel times in the pilot road network. In this case the forecasted data were aggregated in 24-hour intervals (00:00 – 00:59, 01:00–01:59, etc.) for a time-of-day dynamic toll pricing, but the data can also be aggregated in other time granularities to allow for a day-to-day variability or seasonality, or even for real-time calculation.

The final step was to model and validate the dynamic charging strategy. This was accomplished through several iterations with the freight and road infrastructure operators, including two real-world pilot test iterations. The charging strategy is crucial since levying the toll too high may result in underutilization of the tolled highway, while levying the toll too low may lead to excessive migration of vehicles from the free national roads to the highways, leading to the deterioration of the highways' level of service or even lowering the revenue of the road infrastructure operator below the expected affordability values.

To shift the appropriate flow from the national road to the tolled highway, the proposed charging strategy estimates the number of vehicles who are willing to pay a toll for the given travel time savings (i.e., the proportion of freight forwarders/truck drivers whose VoTs are greater than the toll price for a given travel time savings at time t). The number of vehicles on the toll road, Q , should not exceed the available capacity of the toll road, Q_M . Hence, the main variables to fulfill the dynamic toll charging model, besides the VoT, are the difference in travel time between the toll-free and the tolled sections, the current traffic flow (number of vehicles) on the tolled highway section and the maximum number of vehicles allowed on the tolled highway section to preserve the desired Level of Service.

The first three equations (Equations 6.1) are applied if the flow in the tolled section does not exceed the specific threshold provided by the road operator ($Q < Q_M$), related to the minimum LoS that the road operator is obliged to offer. The latter three equations (Equations 6.2) are applied when the flow in the tolled section exceeds the specific threshold, Q_M ($Q > Q_M$). The Portuguese road infrastructure operator defined Q_M , i.e. the maximum vehicle capacity for highways, at 2400 vehicles per hour, since for this pilot, the tolls will be calculated in one hour intervals. This temporal context may be adjusted depending on the specific needs of the system (e.g. tolls changing in real-time, every five minutes, etc.).

$$\forall Q < Q_M \rightarrow T^* = \begin{cases} T & , dt \leq 0 \\ \left[\frac{1}{0,5 + (VoT * dt/400)^{1/4} * (Q/Q_M)^{1/r}} \right] * T & , dt > 0 \end{cases} \quad (6.1)$$

$$\forall Q > Q_M \rightarrow T^* = \begin{cases} T & , dt \leq 0 \\ \left[1 + \left(\frac{dt}{20} \right) \right] * T & , 0 < dt \leq 10 \\ 1,5 * T & , dt > 10 \end{cases} \quad (6.2)$$

Where,

Q	Traffic flow in the tolled section (number of vehicles) for the desired time of toll
Q_M	Maximum traffic flow in the tolled section to ensure the minimum LoS
T^*	The dynamically calculated price
T	The static toll price for the tolled section (charged nowadays)
VoT	The Value of Time (49,4€)
dt	The travel time difference between the national road alternative and the tolled section
r	A parameter that could range from 8 to 50 ensuring that the toll operator has control over the toll price ranges that will be estimated

The model workings are as follows: While the LoS is guaranteed for the tolled section ($\forall Q < Q_M$), and if the travel time in the highway section is shorter than the travel time in the alternative, toll-free national road section ($dt > 0$), then the second member of Equation 6.1 prevails, meaning that the toll price will have a discount given by $\left[1 / [0,5 + (VoT * dt/400)^{1/4} * (Q/Q_M)^{1/r}] \right]$. This member of Equation 6.1 entails that the more traffic there is in the national road alternative, i.e. the bigger the dt , the lower the toll price will be. The main idea is that, if there is too much traffic in the national alternative, then the toll prices are lowered, in order for the traffic to flow to the highway, while the LoS on the highway is guaranteed.

Figure 3 shows the result of the second member of Equation 6.1 on a tolled section in A28 highway (orange), considering the traffic flow in its alternative (blue), for $r = 18$. Parameter r ensures that the toll operator has full control over the toll ranges that will be estimated; a lower r will result in lower toll prices, while higher r will result in higher toll prices. If, however, the travel time in the highway section is bigger than or equal to the travel time in the alternative national road section ($dt \leq 0$), even within the LoS required, then the price of the toll remains static, as represented by the first member of Equation 6.1.

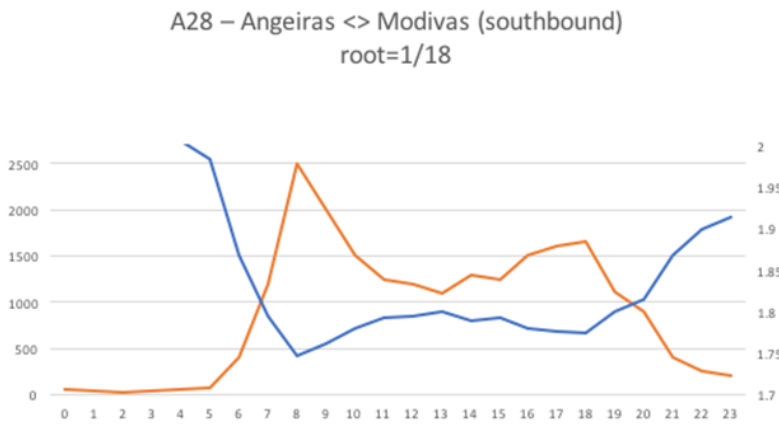


Figure 3. National alternative traffic flow (orange) vs. dynamic toll price in the highway (blue)

When the LoS on the highway degrades to levels below the required threshold ($Q > Q_M$), then Equation 6.2 is used. The first member, as in Equation 6.1, dictates that if the travel time on the tolled highway section is bigger than or equal to the travel time on the alternative, toll-free road ($dt \leq 0$), the toll price remains static, since drivers will not want to go through the slower path, in this case the highway. On the contrary, if the travel time in the highway is shorter than in the alternative ($dt > 0$), the toll price will rise according to $(1 + dt/20)$, until a maximum of 150% of the value of the static toll price, as shown in the second and third members of Equation 6.2. The maximum toll price is enforced when the LoS of the highway is bad, and the travel time on the highway is at least 10 minutes shorter than in the alternative road ($dt > 10$).

5. Big Data support for the Dynamic Toll Charging System

In order to be purely responsive, the OPTIMUM dynamic toll charging system must be able to feed the dynamic charging model with forecasted traffic conditions quickly and efficiently. Time is of the essence when it comes to collect, clean, harmonize, process and apply forecasting or any other analytics techniques on traffic data, to feed it to the model in almost real-time intervals. Hence, the dynamic toll charging system is supported by a Big Data platform, composed by novel technologies that are designed to handle large volumes of heterogeneous data, received at high speeds, sometimes in less than one second. The platform is divided into two main parts, as shown in Section 3: the Observe layer, in which data collection, harmonization and storage is carried out, and the Orient layer, in which processing and analytics tasks are performed over collected data.

5.1. OPTIMUM Data Collection Pipeline

Data collection and harmonization tasks performed in OPTIMUM and their outcomes are described in detail in several scientific publications and conference proceedings (Guerreiro, Figueiras, Silva, Costa, & Jardim-Gonçalves, 2016), (Figueiras, et al., 2016), (Figueiras, Guerreiro, Silva, Costa, & Jardim-Gonçalves, 2018). The OPTIMUM's Data Processing Pipeline is conceptually represented in Figure 4. The pipeline relies on Apache Spark and Apache Storm (The Apache Software Foundation, 2018) Big Data technologies in order to optimize the performance of the data collection and harmonization tasks.

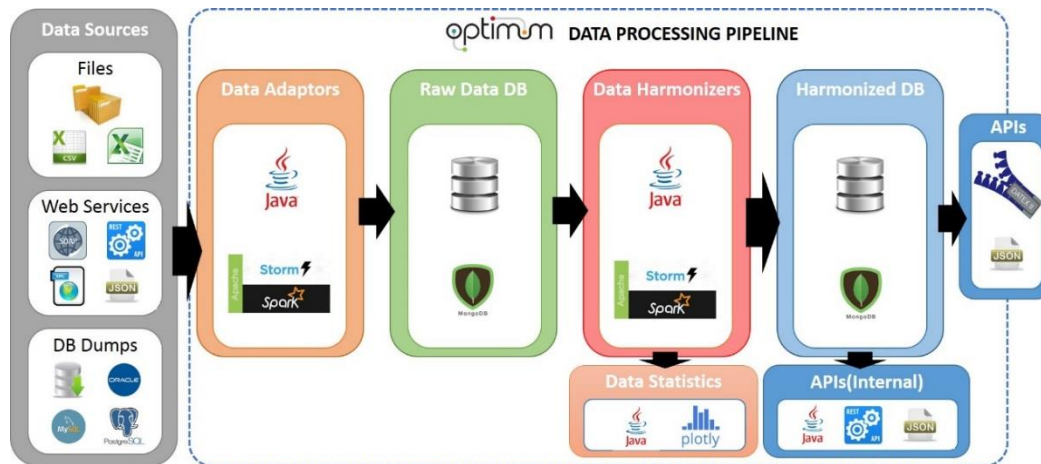


Figure 4. OPTIMUM Data Processing Pipeline

The pipeline's objective is to collect, harmonize and store all data sources linked to the project (e.g. data coming from partners, publicly available data, etc.) and from third-party sources. In order to correctly harmonize all data sources into generic formats, a conceptual analysis of all mobility-related data under the OPTIMUM project was performed. This analysis resulted in the definition of standardized data formats for mobility information, under the scope of the OPTIMUM project. These harmonized schemas for each type of data (weather, traffic sensors, events, bike sharing, car parking, etc.) were created in a way that all data sources are from the same type, but come from different data providers, in different formats, which could be stored in a single harmonized format per data type, which all OPTIMUM services could use as inputs. The schema is composed by a JSON-based meta-schema, meaning that if some new data format is needed, the system just needs the JSON meta-schema for that specific data format, in order for it to be compliant with the OPTIMUM data collection pipeline.

Data adaptors were then developed for the different data sources, depending on format (XML, JSON, CSV, database dumps, etc.) and type. The pipeline is generic and ready to integrate new adaptors, since it enables the addition of other adaptors that implement the adaptor interface. Hence, when new data sources are provided, new adaptors may be aggregated to the pipeline to handle these new data sources. These adaptors

collect data and store it in an intermediate raw data database, which prevents any kind of data loss or corruption.

Next, generic data harmonizers were developed in order to harmonize raw data into OPTIMUM's harmonized formats. Harmonizers are also generic in the sense that they accept a JSON schema as input that represents the raw data format, meaning that for each data source a different schema is provided to the harmonizer, without the need for having one specific harmonizer for each data source. On the other hand, OPTIMUM's harmonized schemas are also represented in JSON, allowing for new harmonized schemas to be added later to the pipeline, for new data types. The harmonizer gets the raw data and maps its fields to the harmonized schema's fields. When the harmonization is done, harmonized data can be stored in the main database, as in the case of batch offline data, or be served directly to other layers of the system (e.g. Orient layer), as in the case of real time streaming data.

Several Application Programming Interfaces (APIs) were implemented in order to provide access to harmonized data to the next OODA phases, or for use in external, third-party applications. These APIs serve data in two specific formats: DATEX II (EIP/EIP+ Project, 2014), the European *de facto* data representation standard for ITS applications and services, and JSON. For real time streams of data, JSON is used to pass data chunks, or messages, to the higher layers of the architecture and to their services, such as the forecasting and analytics processes.

Also, statistical analyses are applied to the harmonized data so as to have an overview of the data being stored. These analyses consist in, for instance, counting the number of collected records, presenting the raw data schema and a visual representation of a sample of the raw data, as well as performing outlier detection and calculating some figures, such as means or standard deviations. For an overview of the performance of the pipeline, please refer to (Figueiras, Guerreiro, Silva, Costa, & Jardim-Gonçalves, 2018).

The pipeline has a built-in Web Application, with a User Interface for third-party companies to upload their data into OPTIMUM. Third-party company users are invited to point out the paths to their data sources, and to provide the raw data schema. After this, the process begins, as previously described. When the harmonization process finishes, two options are available: a data visualization tool, in which statistics are shown for the collected data, and a data export tool, which can be used to export collected data in DATEX II or JSON.

5.2. OPTIMUM Big Data Forecasting Architecture

OPTIMUM's Traffic Forecasting service (Petalas, Ammari, Georgakis, & Nwagboso, 2017) is shown in Figure 5. The presented architecture relies on several Big Data technologies (Apache's Spark, Flume and Kafka (The Apache Software Foundation, 2018)) and novel data storage technologies (Apache's HBase (The Apache Software Foundation, 2018) and Redis (RedisLabs, 2017)) to enable performance-optimized forecasting processes. This distributed architecture was developed to efficiently support transport related predictive analytics services that use information from heterogeneous sources such as traffic sensors, social media, weather stations and others.

The data input streams, coming from the Data Collection Pipeline, are received in Apache Kafka. Apache Kafka's main characteristics are persistence and replication messaging, high throughput, low latency, fast real-time consumption of messages, high availability and no loss of data. With the usage of a pub-sub system, like Kafka, the decoupling of the data sources from the other parts of the architecture is achieved, thus additional data sources can be easily integrated. This approach allows the expandability of the architecture to support additional transportation services in the future. Data from Kafka is then transferred to Hadoop (The Apache Software Foundation, 2018), where it can be stored for lengthy periods of time. The volume of data per year can amount to terabytes, thus the usage of the Hadoop ecosystem is adopted.

Data reaches Hadoop through Apache Flume, which can store it to HDFS or to any other storage technology supported by Hadoop. For each topic created in Kafka an agent in Flume is created to transfer the data to a

table in HBase. In the terminology of Flume, the source is the Kafka topic, and the sink is the table in HBase.

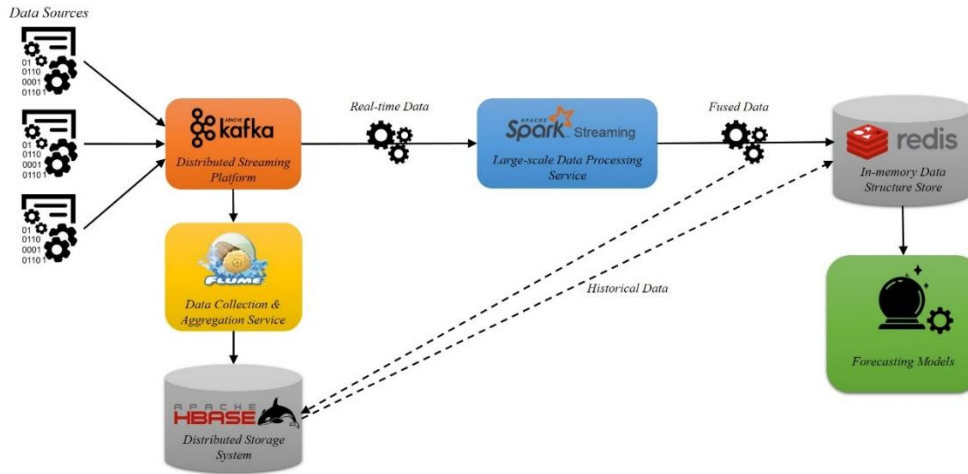


Figure 5. Big Data forecasting architecture overview, based on (Petalas, Ammari, Georgakis, & Nwagboso, 2017)

A record in HBase is composed of a row-key and column families. Each column family contains columns for the available data, and each record will only have columns for the available measurements from a source (e.g. traffic sensor). For the traffic data the row-key has the format *sensorId_datetime*, (*datetime* in format of milliseconds) and there is one column family (“traffic”) with a column for each of the available traffic measurements. For example, if a sensor has values for flow and speed, there will be two columns in this record. Equivalently, if a sensor has values for flow, speed, headway and travel time, there will be four columns.

Data from the different sources is received in different granularities and presents different levels of quality. Therefore a ‘Data Fusion’ component has been developed as part of the architecture. This component creates fused datasets that contain information merged from all the available data sources (e.g. traffic measurements, weather or social media data). In this procedure, data from the available sources for a specific day, contained in HBase tables, is retrieved. A procedure for cleaning is performed, taking into account the number of measurements per day for each sensor, the number of zeros and the range of the values (e.g. speed values could not be higher than 120km/h in an urban area).

The traffic data is aggregated at the specified time granularity (e.g. hourly) and the missing values are filled using linear interpolation. Spark Streaming and a Spark-based library for time series is used for the resampling (down-sampling, up-sampling) and the filling of missing values. Furthermore, traffic data from different sources have different requirements in terms of aggregation and cleaning, so custom filtering and cleaning tasks are performed before training the forecasting models, in order to enable the exclusion, or dampening of the effect that erroneous, or outlier, data have on the forecasting output.

This fused dataset from Spark Streaming is stored in a memory database, Redis. The role of the in-memory database is to serve as a cache, storing the most recent data, thus allowing for rapid retrieval and real-time processing. Redis has data structures such as sorted sets and hashes, for handling efficiently time series and a connector is available for communicating with Spark.

5.2.1. Reducing computational time on training forecasting models

The forecasting models were trained offline from the data that was stored in the aggregated/fused tables in HBase. Even if there is data of several years and from many sensors, the training for single point forecasting is performed per sensor and for a single traffic measurement. Data is retrieved from HBase per traffic measurement (e.g. traffic flow) and for a specific time period, (e.g. the period 1 May 2017- 1 June 2017), and

passed to a Spark's Resilient Distributed Dataset (RDD). This RDD contains all the sensors that have traffic measurements in this period. For each row of this RDD, a function is applied and thus for each sensor, training algorithms are executed. Hadoop's cluster manager and scheduler decides, based on optimal processing balancing, in which of the cluster's nodes the forecasting models will be executed.

A practical issue that arises from the use of data from different sensors in various locations of a transportation network is the generation of a predictive model for each sensor within the area of interest. In such cases the creation of a forecasting model for each sensor by getting the necessary data (traffic measurements) to train the model, and the regular and quick updates of this model, as new data is generated, requires a lot of computational resources. Hence, one possible solution for this issue is to form clusters by grouping sensors with similar characteristics. To accomplish this task, clustering techniques related to time series can be used.

A model-based approach for the clustering task was employed. Specifically, the methodology presented by (Kalpakis, Gada, & Puttagunta, 2001) was adopted, along with an auto-regressive (AR(2)) model, which was fitted for traffic flow time series of each sensor and from the coefficients of the fitted model 10 *cepstrum* coefficients were computed and used as the feature vector of each sensor for the clustering. From preliminary experiments, the AR(2) model was the best model for most of the sensors. The clustering algorithm used was the k-medoids algorithm, and Euclidean distances were used to calculate distances between objects.

The second step was the evaluation of the clustering results and the determination of the number of clusters. The Silhouette coefficient was used for the evaluation of the clustering procedure. It is calculated using the average distance between an object and the other objects in the same cluster and between an object and the objects in the other clusters. Hence, for a group of 120 sensors, the Silhouette coefficient had the highest value 0.56 for $k=8$, thus 8 clusters were chosen having 3,32,24,9,7,11,21,13 objects (sensors) respectively. Finally, after the application and evaluation of the clustering algorithm, the most representative sensor for each cluster was chosen, and a forecasting model for that sensor was created. The centroid of each cluster, the object with the minimum average distance from all the other objects in the cluster, was chosen.

5.2.2. Experiments

To test the efficiency and performance of the developed architecture, a number of experiments, using different forecasting models, was performed. For the purposes of these experiments, one month of traffic flow data for the 120 sensors, between middle of January and middle of February 2017, were used and these were aggregated per hour. Four forecasting models were implemented, namely, Linear Regression, Support Vector Regression (SVR), Artificial Neural Networks (ANNs) and Random Forest Regression. 80% of the available traffic data for each sensor was used as training set and the rest 20% as test set. The evaluation metric used was the Mean Absolute Percentage Error (MAPE), since it is commonly used in the literature for traffic forecasting.

In the experiments, the forecasting models predicted traffic flows for 12 steps ahead (ie. for the next 12 hours). For each methodology, except ANNs, 12 models were used, one for each forecasting step. This method produced better results than the incremental approach for multistep ahead prediction, where the output from a predicted step is fed back as input to provide the next prediction value. For ANNs, one model was used, having 12 neurons as outputs, each of one representing a number from the prediction steps ahead. An evaluation of the proposed methodology has been performed by comparing the prediction errors for each sensor, having a forecasting model created for each one, and the predictions for the same sensor using the forecasting model produced for the medoid of the cluster it belongs to. For this reason, for each sensor the difference was computed between the prediction error (MAPE) before the application of the proposed approach and afterwards. This took place for all the 4 forecasting models and the 12 prediction steps ahead (Table 2).

For Linear Regression and Random Forest models, the average value is close to zero and the standard deviation is relative small. This indicates that most of the values are around the average value. In the case of the SVR model, there are cases that the proposed methodology has worst prediction errors but in average, and

as in the cases presented above, the average differences approximate zero and have small standard deviations. Finally, ANNs demonstrate small negative values and in some cases there is significant reduction in the prediction error reaching step 7. The average differences and standard deviations follow the same trends as those in the aforementioned methods.

Table 2. Prediction mean errors and standard deviations (MAPE) for the four forecasting models tested

<i>Model</i>	Linear Regression		Random Forest		SVR	ANN		
<i>Step</i>	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
1	0.02	0.24	-0.31	0.53	-0.22	0.64	-0.03	0.36
2	0.01	0.39	-0.39	0.52	-0.30	0.50	-0.08	0.35
3	-0.01	0.55	-0.48	0.48	-0.26	0.62	-0.04	0.39
4	-0.03	0.71	-0.51	0.56	-0.33	0.56	-0.03	0.43
5	-0.06	0.86	-0.67	0.66	-0.37	0.74	-0.01	0.53
6	-0.07	1.00	-0.62	0.81	-0.34	0.60	-0.03	0.62
7	-0.09	1.15	-0.94	0.88	-0.30	0.57	-0.05	0.61
8	-0.12	1.31	-0.95	1.12	-0.31	0.51	-0.01	0.63
9	-0.15	1.48	-0.99	1.18	-0.34	0.58	-0.02	0.84
10	-0.19	1.64	-0.98	1.30	-0.38	0.59	-0.01	0.77
11	-0.19	1.81	-1.19	1.47	-0.43	0.75	0.01	0.80
12	-0.18	1.99	-1.18	1.57	-0.34	0.81	-0.03	0.84

Unfortunately, in statistical learning, there is the concept of “no free lunch”. There is not one method that will be the best for all the problems. Thus, in literature, a combination of models has been proposed to overcome the issue of having to select a single model for each problem under study. Using the combinations of the predictions of a set of models one final, combined, prediction is produced. For details on the proposed methodology, please refer to (Petalas, Ammari, Georgakis, & Nwagboso, 2017). In this study, three approaches for the combination of forecasting models were developed. The first uses the minimum value from the individual models as the final prediction (MIN). The second was Stacked Regression, along with another approach based again on Stacked Generalization using an Artificial Neural Network to receive the inputs of the individual forecasting models and producing as output the final prediction, called Stacked Neural.

Ten-fold cross validation and least squares optimization methodologies were used to solve the optimization problem and calculate the coefficients for Stacked Regression. The dataset consisted from traffic flow from 30 sensors during the period 21 June 2017 – 21 August 2017. The data was aggregated hourly and the forecasting horizon was 12 hours ahead (12 steps). The absolute differences between the prediction error from the best individual model and the models from the proposed methodology were computed for each sensor along with some statistics of the differences for all the sensors used in the experiments. For exemplification, Table 3 shows, for a single sensor, the best and the worst model in terms of forecasting accuracy are presented. It is obvious that there is not a single model that will perform better in all the cases.

Table 3. Best and worst model per forecasting step (single sensor)

Step	Single Best model	Single worst model
1	Forest	SVR
2	Neural	Forest
3	Linear	Forest
4	Linear	Forest
5	Neural	Forest
6	Neural	SVR
7	Forest	SVR
8	Linear	Neural
9	Forest	SVR
10	Forest	SVR
11	Linear	Neural
12	Forest	SVR

In Table 4, the median MAPE errors for all the employed methods per forecasting step are shown. Practically, the best individual model and the proposed combined forecasting models, (MIN, Stacked Regression, Stacked Neural) achieved the same prediction accuracy.

Table 4. Median MAPE of all methods per forecasting step

Step	Best individual	MIN	Stacked Regression	Stacked Neural
1	9.85	9.98	11.00	10.96
2	11.45	11.35	12.00	11.69
3	11.94	11.68	12.53	12.01
4	12.38	12.17	12.32	12.18
5	12.76	12.45	12.79	12.51
6	13.09	12.77	12.87	12.89
7	13.10	12.83	12.80	12.86
8	13.37	13.22	13.29	13.27
9	13.56	13.44	13.75	13.42
10	13.66	13.62	13.69	13.71
11	13.64	13.78	13.68	13.63
12	13.59	13.35	13.84	13.29

Several statistical tests were performed to investigate if the proposed methods had the same forecasting accuracy as the best individual model. MIN had the same forecasting accuracy as the best individual model in all the forecasting steps and statistical tests. Both stacked methods achieved the same forecasting accuracy as the best individual model after forecasting step 6. Thus, the proposed methodologies can be used with the combination of forecasting models to achieve the same accuracy as the best individual model in each case, which of course is not known in advance.

6. Validation and Discussion

This section describes the evaluation methodology and results from OPTIMUM dynamic toll charging system. The two end-users of the system were LS, the freight operator and IP, the road operator. The pilot business processes are overviewed in Figure 6, a). After the definition of an agreement between IP and LS, and the subscription of LS to the Dynamic Toll Pricing service managed by IP, the toll prices are calculated, taking into account the forecasted traffic flows in both the tolled highway and the toll-free alternative, and published in the service. LS then plans the freight routes considering the dynamic toll prices, and effectively crosses the toll sections. The toll system recognizes LS vehicles' electronic toll charging devices and the toll rate with discount is applied by the concessionaire, following the normal procedure of the toll system payment.

Since LS needs to plan routes two days in advance, in order to manage the freight vehicles operation, usage and availability, toll prices were provided with a 48-hour span ahead of time, based on forecasted traffic flows for the highways, allowing them to plan their routes in advance. Both operators were requested to store and provide data for the duration of the pilot. From LS, data regarding kilometers driven, time spent, operated routes, fuel costs and total costs, before (baseline) and during the pilots, were obtained from the 10 trucks that were used in the pilot, while data on toll revenues were provided from IP.

Figure 6, b) depicts the pilot's key performance indicators (KPIs), already presented in Section 1.2. Specifically, the main contributions of the pilot should be a 15% freight traffic shift from urban and national roads to tolled highways, which would allow for a 10% increase on IP's toll revenues and also a 10% improvement in the efficiency of LS' operations.

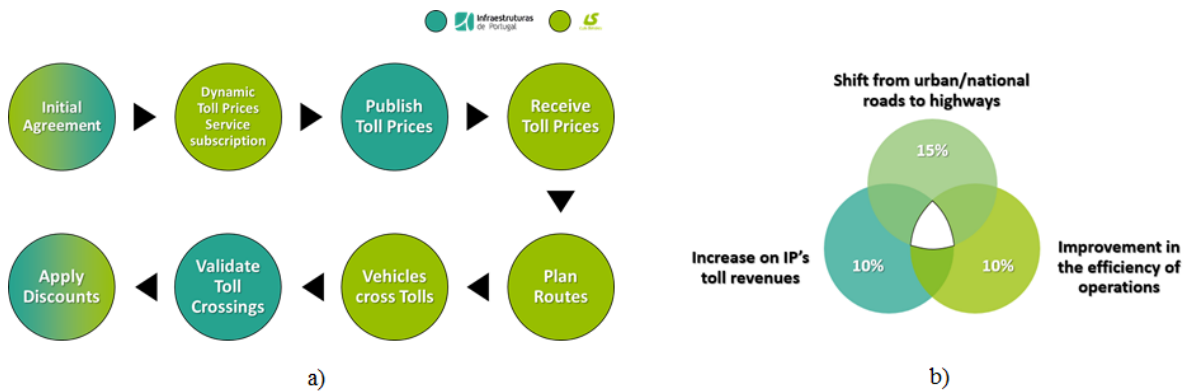


Figure 6. The OPTIMUM Portuguese Pilot. a) Steps; b) KPIs

The pilot used 10 trucks from LS, and was divided into two iterations. The first iteration enabled the validation of the initial version of the dynamic charging model, which provided useful insights on the model's workings and what had to be changed, and also the testing of the first version of the Big Data platform that supports the system as a whole. The second iteration enabled the demonstration of the final version of the dynamic toll pricing strategy and system and the validation of the pilot's KPIs. This section mainly evaluates the second iteration of the dynamic charging pilot.

6.1. Traffic shift from urban/national roads to highways

Shifting traffic from the national roads to the highways is the main goal of this pilot. IP has a high cost with the national road network maintenance. This cost could be significantly lower if the national road traffic was reduced, as they were not projected for such high levels of traffic especially from heavy vehicles. A comparison between the Optimum routes and the baseline (original routes) was achieved, using GPS data collected by LS, i.e. the kilometers driven by the trucks in each trip, in national roads and highways, and is presented in Table 5.

Table 5. Traffic shift KPI

	Total	National Roads		Tolled Highways	
	<i>Km</i>	<i>Km</i>	%	<i>Km</i>	%
LS Original Routes (baseline)	46,612	16,250	35%	30,362	65%
OPTIMUM Routes	46,208	11,945	26%	34,263	74%
Variation	-0,9%	-26%	-	+13%	-

With the use of the dynamic charging model, a reduction on the use of the national road (of about 26%) was observed. This rate is significantly higher than the target defined for this KPI (15%). The pilot also led to an increase in the use of highways (of 13%) in terms of kilometers driven. Thus, the share of the highway's traffic increased to 74% (in comparison to 65% before the pilot). The detailed analysis of the routes chosen by LS revealed that in many cases there was a traffic shift from other highways, mainly private concessions, which leads to better results for IP, in terms of traffic and revenues.

6.2. Cost improvement in the efficiency of LS operations

Table 6 presents a summary of LS operations before and during the pilot. The efficiency of LS's operations cost increased in a substantial way, precisely by 5.48%. This change is attributed to two main factors: a) a reduction of 10% in fuel consumption was observed, which is important from the freight operator's perspective.

Table 6. Improvement of LS operations KPI

	Baseline	OPTIMUM	Variation
Number of routes	422	422	-
Total Kms	46,612	46,208	-0,87%
Total time spent on trip (hours)	758:18:47	711:29:00	-6%
Average time spent per trip (hours)	1:47:49	1:41:10	-6%
Total cost (€)	50.241,23	47.487,25	-5,48%
Average cost per trip (€)	119,06	112,53	-5,48%

This could be explained by the fact that LS chose highways instead of national roads in many situations and benefited from highways' better conditions of traffic and steady speed; 2) the second factor is related to the reduction of travel time of the trips by 6%. This fact helped drivers to do more trips or reduce the risk of getting late to destination.

6.3. Increase on IP's toll revenues (heavy vehicles)

IP's toll revenues were calculated by, once more, comparing the baseline routes with the OPTIMUM routes. The pilot had a very good performance regarding this KPI, as a 14% increase in revenues exceeded the defined goal of 10%, as shown in Table 7. These results refer only to the 10 LS trucks operating in the pilot. Nevertheless, an extrapolation of these results with regards to the number of heavy freight vehicles operating in the Portuguese road network provides crucial evidences of the impact of the dynamic charging system on IP's toll revenues.

Table 7. IP toll revenues KPI

IP toll revenues from baseline routes	IP toll revenues from OPTIMUM routes	Variation
3,649€	4,169€	+14%

More specifically, in 2016, the four highway concessions that formed the pilot generated revenue of nearly 30 million euros for heavy vehicles (class 4). If the result of this KPI is extrapolated to the total number of heavy freight vehicles in these concessions, even considering only an increase of 10% of the revenues, this would result in an increase of 3 million euros in yearly toll revenues.

6.4. Discussion and Conclusions

Regarding the quantitative evaluation, Table 8 provides the comparison of the actual results from the pilot against the KPIs that were set at the beginning. As indicated, the traffic shift and the revenues KPIs were

successfully reached. Specifically, the dynamic charging pilot resulted in 26% reduction in the number of kilometers driven on national roads (overcoming the target KPI of 15%) and 14% increase in IP's toll revenues (compared with the 10% target). Regarding the cost improvement KPI, the overall travel time of LS trucks decreased by 6%, while the respective fuel consumption decreased by 10%. All in all, LS was able to achieve the target of cost improvement through the use of the dynamic charging model.

Table 8. Comparison of pilot's results with initial KPIs

	KPI Description	Target	Pilot Results
1	Traffic shift (Reduction of the number of kilometers driven in national roads)	-15%	-26%
2	Operational improvement (Reduction in travel times and fuel consumption)	-10%	-6% travel time -10% fuel consumption
3	Revenues (Increase of IP's toll revenues)	+10%	+14%

The implementation of the dynamic charging model over-performs the static pricing in at least three categories: (i) it shifts traffic from the urban/national road network to the underused toll network (a 26% reduction in the number of kilometers driven in the national roads was observed during the pilot), (ii) it optimizes the freight operators cost and efficiency of operations and fulfilling the main goal of the pilot (a 6% reduction in LS's travel time and a 10% reduction in fuel consumption were observed) and (iii) it yields more revenue for the toll operator in the category of heavy vehicles (a 14% increase in IP's toll revenues was observed).

As far as the model transferability is concerned, the modelling framework allows for a certain level of customization in order to satisfy the needs of the stakeholders in the freight transport industry. This customization is available at the macro or micro location level, with different values of time for countries and regions and different parameters of the model for specific road sections. The system is also customizable to either achieve maximum profit or ensure the optimal usage of the traffic flow through the road network. The findings from the implementation in Portugal reveal that the dynamic charging system has higher performance and yields a more substantial net revenue, outperforming the static toll system in practice today. The dynamic charging model developed is focused on shifting traffic from national roads to toll roads. The proposed methodology can be adjusted according to the needs and objectives of the partners and stakeholders. For example, the methodology could be adjusted to optimize the toll prices for maximum profit of the road operator.

Both the road operator, IP, and the logistic operator, LS, expressed their overall satisfaction with the pilot results via in-depth interviews conducted with two high representatives of both operators, after the pilot. From IP's perspective, the Big Data-supported system's functionality and performance, as well as its responsiveness were indicated as being extremely satisfying. The ease of use and customization of the dynamic charging strategy was also regarded as a really important factor. The existence of random events (like unexpected delays or accidents) is referred by the road operator as potential triggers to change the prices in an effort to manage the available road capacity. Planned events, like roadworks or national holidays, are also possibilities to be explored as factors affecting the toll price and the model functions.

For LS, the Dynamic Toll Charging system seems useful and has significantly affected the company's routing choices. Regarding the strongest points of the system, LS stated that these were the 48 hour-ahead vision and the possibility of real time calculation of discounts based on forecasted traffic conditions. In addition, in a very competitive sector as the logistic sector, with historical low profit rates, the reduction of the operation cost by over 5% was very important. Furthermore, it was stated that the dynamic charging scheme implemented in this pilot provided insights that really helped LS to change the traditional way of using highways.

Hence, the overall satisfaction with the system, among the main stakeholders, was very good, and the initial objectives were fully accomplished. Moreover, the novel characteristics of both charging scheme and Big Data support were proven very effective when working together, since they present a symbiotic relationship: On one hand, for the dynamic charging scheme to achieve its goals, more detailed information is needed (e.g. traffic flows, traffic events, weather, etc.); on the other hand, dealing with big amounts of heterogeneous data, by cleaning, processing and applying forecasting algorithms on it, will entail that the algorithm will be more refined.

Nevertheless, the fact that the system enabled a shift of freight traffic from national roads to tolled highways does not mean that the average driver would also do it. Furthermore, and since freight operators plan their trips in advance, it was easier to present the toll prices based on forecasted traffic conditions to the drivers via a Web interface, because routes were calculated with two day of advance, in some cases. A more efficient and swift way of persuading drivers to shift to dynamically tolled highways is necessary, such as on-board mobile notifications, for instance, alerting drivers not only about traffic congestions on the road ahead, but also how can they divert from them and, at the same time, pay less tolls by doing it. Even so, due to law regulations, which state that the tolls are fixed by law, the shift from static to dynamic tolls has yet some time to mature in research endeavors such as the OPTIMUM project, and one plausible way to mature is to validate the proposed system with a bigger slice of the population, especially, with regular drivers.

References

- Addo-Tenkorang, R., & Helo, P. T. (2016). Big data applications in operations/supply-chain management: A literature review. *Computers & Industrial Engineering*, 101(November 2016), 528-543.
- Audretsch, D. B., Dohse, D. C., & dos Santos, J. P. (2017). Do Toll-free Highways Foster Firm Formation and Employment Growth? Results from a Quasi-natural Experiment. *KIEL Working Paper*, 2080. Obtido de https://www.ifw-kiel.de/fileadmin/Dateiverwaltung/IfW-Publications/Dirk_Christian_Dohse/do-toll-free-highways-foster-firm-formation-and-employment-growth-results-from-a-quasi-natural-experiment/kwp_2080.pdf
- Bandeira, J., Coelho, M., Pimentel, M., & Khattak, A. (2012). Impact of intercity tolls in Portugal - An environmental perspective. *Transport Research Arena - Europe*. Athens, Greece.
- Beckmann, M. (1965). On Optimal Tolls For Highways, Tunnels, and Bridges. Em *Vehicular Traffic Science* (pp. 331-341). New York, USA: American Elsevier Publishing Company.
- Ben-Akiva, M., & Lerman, S. R. (1988). Discrete Choice Analysis: Theory and Applications to Travel Demand. *Journal of Business & Economic Statistics*, 6(2), 286.
- Bracher, B., & Bogenberger, K. (2017). A dynamic pricing scheme for a congestion charging zone based on a network fundamental diagram. *5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems*. Naples, Italy.
- Button, K. J. (1993). *Transport Economics* (2 ed.). London, England: Edward Elgar Publishing Ltd.
- Cheng, D., & Ishak, S. (2016). Maximizing Toll Revenue and level of service on managed lanes with a dynamic feedback-control toll pricing strategy. *Canadian Journal of Civil Engineering*, 43(1), 18-27.
- de Palma, A., & Lindsey, R. (2011). Traffic congestion pricing methodologies and technologies. *Transportation Research Part C*, 19, 1377-1399.
- Deck, C., & Kimbrough, E. O. (2014). Paying for Express Checkout: Competition and Price Discrimination in Multi-Server Queuing Systems. *Public Library of Science*, 33.
- Dinis, D., Barbosa-Póvoa, A., & Teixeira, Â. P. (2018). Valuing data in aircraft maintenance through big data analytics: A probabilistic approach for capacity planning using Bayesian networks. *Computers & Industrial Engineering*, In Press.
- EIP/EIP+ Project. (2014). *Datex Easyway*. Obtido de DATEX support web site: <http://www.datex2.eu/>
- European Commission. (2003). *5th RTD Framework Programme*. Belgium.
- European Union. (2010). *Intelligent Transport Systems: EU-funded research for efficient, clean and safe road transport*. European Commission, Directorate-General for Research Transport. Belgium: Publications Office of the European Union. Obtido de [http://www.gppq.fct.pt/h2020/_docs/brochuras/transportes/intelligent transport systems.pdf](http://www.gppq.fct.pt/h2020/_docs/brochuras/transportes/intelligent%20transport%20systems.pdf)
- European Union. (2011). DIRECTIVE 2011/76/EU OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL. *Official Journal of the European Union*(amending Directive 1999/62/EC on the charging of heavy goods vehicles for the use of certain infrastructures).
- Fahmideh, M., & Beydoun, G. (2018). Big data analytics architecture design—An application in manufacturing systems. *Computers & Industrial Engineering*, In Press.
- Fang, K., Jiang, Y., & Song, M. (2016). Customer profitability forecasting using Big Data analytics: A case study of the insurance industry. *Computers & Industrial Engineering*, 101(November 2016), 554-564.
- Federal Highway Administration. (2017). *Congestion Pricing — A Primer*. U.S. Department of Transportation.
- Figueiras, P., Guerreiro, G., Silva, R., Costa, R., & Jardim-Gonçalves, R. (2018). Data Processing and Harmonization for Intelligent Transportation Systems: An Application Scenario on Highway Traffic Flows. Em V. Sigurev, V. Piuri, & V. Jotsov (Edits.), *Learning Systems: From Theory to Practice* (pp. 281-301). Springer.
- Figueiras, P., Silva, R., Ramos, A., Guerreiro, G., Costa, R., & Jardim-Goncalves, R. (2016). Big Data Processing and Storage Framework for ITS: A Case Study on Dynamic Tolling. *ASME 2016 International Mechanical Engineering Congress and Exposition*. Phoenix, AZ, USA. doi:10.1115/IMECE2016-68069
- Governo de Portugal. (26 de October de 2012). Portaria n.º 342/2012 of 26 of October. *Diário da República*, pp. 6049-6055. Obtido de <https://dre.pt/application/file/192497>

- Governo de Portugal. (20 de July de 2016). Portaria n.º 196/2016 of the 20 of July. *Diário da República*, pp. 2365-2370. Obtido de <https://dre.pt/application/file/a/74981761>
- Guerreiro, G., Figueiras, P., Silva, R., Costa, R., & Jardim-Gonçalves, R. (2016). An architecture for Big Data processing on Intelligent Transportation Systems: An application scenario on highway traffic flows. *IEEE Intelligent Systems 2016*. Sofia, Bulgaria. Obtido em December de 2016
- Gutman, P.-O. (2016). Dynamic Pricing for Toll Lanes - a Case Study. *IFAC OL 2016*.
- Hourdos, J., Janson, M., Levinson, D., & Parikh, G. (2015). *MnPASS Modeling and Pricing Algorithm Enhancement*. St. Paul, Minnesota: Minnesota Traffic Observatory, Department of Civil, Environmental, and Geo-Engineering, University of Minnesota. Obtido de <http://www.dot.state.mn.us/research/TS/2015/201522.pdf>
- Instituto de Mobilidade e dos Transportes. (2014). *Relatório de Monitorização da Rede Rodoviária Nacional 2012 e 2013*. Lisbon, Portugal: Instituto da Mobilidade e dos Transportes, IP .
- Jang, K., Chung, K., & Yeo, H. (2014). A dynamic pricing strategy for high occupancy toll lanes. *Transportation Research Part A: Policy and Practice*, 67, 69-80.
- Kalpakis, K., Gada, D., & Puttagunta, V. (2001). Distance measures for effective clustering of ARIMA time-series. *International Conference on Data Mining* (pp. 273-280). IEEE.
- Kucharski, R., & Drabicki, A. (2017). Estimating Macroscopic Volume Delay Functions with the Traffic Density Derived from Measured Speeds and Flows. *Journal of Advanced Transportation*, 2017, 10.
- Kumar, G. N., & Machado, H. K. (2018). EATS: Enhanced Automatic Tollgate System Powered by IoT and Big Data. *International Journal of Scientific & Engineering Research*, 9(1), 967-972.
- Lee, C. H., & Yoon, H.-J. (2017). Medical big data: promise and challenges. *Kidney Research and Clinical Practice*, 36(1), 3-11.
- Leonhardt, A., Saches, T., & Busch, F. (2012). Dynamic control of toll fees for optimal high occupancy toll (HOT) lane operation. *91 st Annual Meeting of the Transportation Research Board*. Washington, D.C.
- Li, B., Ch'ng, E., Chong, A. Y.-L., & Bao, H. (2016). Predicting online e-marketplace sales performances: A big data approach. *Computers & Industrial Engineering*, 101(November 2016), 565-571.
- Li, J. (2001). Explaining High-Occupancy-Toll Lane Use. *Transportation Research Part D*, 6, 61-74.
- Li, Y., Saigal, R., & Zhou, H. (2012). Distance-based dynamic pricing strategy for managed toll lanes. *Transportation Research Record Journal of the Transportation Research Board*, 2283, 90-99.
- Liu, J.-W. (2018). Using big data database to construct new GFuzzy text mining and decision algorithm for targeting and classifying customers. *Computers & Industrial Engineering*, In Press.
- Liu, Z., Wang, S., & Meng, Q. (2014). Toll pricing framework under logit-based stochastic user equilibrium constraints. *Journal of Advanced Transportation*, 48(8), 1121-1137.
- Los Angeles County Metropolitan Transportation Authority. (2017). *HOT Lanes in the U.S*. Obtido em 8 de March de 2017, de https://www.metro.net/projects/expresslanes/expresslanes_us/
- Lou, Y., Yin, Y., & Laval, J. (2011). Optimal dynamic pricing strategies for high-occupancy/toll lanes. *Transportation Research Part C*, 19, 64-74.
- McQueen, B. (2017). *Big Data Analytics for Connected Vehicles and Smart Cities*. Boston, MA, USA: Artech House.
- Morgul, E. F. (2010). *Simulation Based Evaluation of Dynamic Congestion Pricing Algorithms and Strategies*. New Jersey, USA: Graduate School – New Brunswick Rutgers, The State university of New Jersey.
- Murray, P. (2012). Congestion Pricing for Roads: An Overview of Current Best Practice, and the Economic and Transport Benefits for Government. *Public Infrastructure Bulletin*, 1(8), Article 8.
- Naga, R. (2007). *A Mathematical Model for Evaluating the Conversion of High Occupancy Vehicle Lane to High Occupancy/ Toll Lane*. UC Davis - Institute of Transportation Studies.
- OECD. (2013). Portugal: reforming the state to promote growth. *"Better Policies" Series*, 1-80.
- OECD. (2015). Road Traffic, vehicles and networks. Em *Environment at a Glance 2015*. OECD.
- OPTIMUM consortium. (1 de October de 2015). *OPTIMUM Project*. Obtido em 20 de April de 2016, de <http://optimumproject.eu/>
- Osinga, F. (2005). *Science, Strategy and War: The Strategic Theory of John Boyd*. Delft, NL: Eburon Academic Publishers.

- Petalas, Y. G., Ammari, A., Georgakis, P., & Nwagboso, C. (2017). A Big Data architecture for traffic forecasting using multi-source information. *International Workshop of Algorithmic Aspects of Cloud Computing*. Aarhus, DK. doi:10.1007/978-3-319-57045-7_5
- Ramos, A., Rodrigues, A., Machado, S., Antunes, F., Ventura, P., Martins, A., & Kioussi, A. V. (2018). Making Big Data Real in Upcoming Future: The Dynamic Toll Prices in the Portuguese Highways. *The 4th Conference on Sustainable Urban Mobility*. Skiathos, Greece.
- RedisLabs. (2017). *Redis*. (RedisLabs) Obtido de <https://redis.io/>
- Santos, M. G., & Santos, B. F. (2012). Shadow-tolls in Portugal: How We Got Here and What Were the Impacts of Introducing Real Tolls. *European Transport Conference 2012*. Glasgow, Scotland.
- Seik, F. T. (1997). An Effective Demand Management Instrument in Urban Transport: The Area Licensing Scheme in Singapore. *Cities*, 14(3), 155-164.
- Shirazi, M., Aashtiani, H. Z., & Quadrifoglio, L. (2017). Estimating the minimal revenue tolls in large-scale roadway networks using the dynamic penalty function method. *Computers & Industrial Engineering*, 107(May 2017), 120-127.
- The Apache Software Foundation. (2018). *Projects List*. Obtido em 15 de 01 de 2019, de The Apache Software Foundation: <https://www.apache.org/index.html#projects-list>
- The Texas A&M Transportation Institute . (2015). *Urban Mobility Scorecard 2015*. The Texas A&M Transportation Institute , The Texas A&M Transportation Institute . Texas: Texas A&M.
- Unidade Técnica de Acompanhamento de Projectos. (2017). *Boletim Anual das PPP – 2016*. Lisbon, Portugal: Ministério das Finanças.
- United Nations. (2018). *Revision of World Urbanization Prospects*. DESA/Population Division.
- Vickrey, W. S. (1969). Congestion Theory and Transport Investment. *American Economic Review*, 59(2), 251-260.
- Walters, A. A. (1961). The Theory and Measurement of Private and Social Cost of Highway Congestion. *Econometrica*, 29(4), 676-699.
- Winaisathaporn, P. (s.d.). *Estimation of a Speed/Density Model of Marginal Congestion Cost and the Economic Cost of Roadworks*. University of Kent, School of Economics. Kent: University of Kent.
- Wise, P. (2013, August 25). High toll charges leave Portugal's drivers on road to nowhere. *Financial Times*. Retrieved from <https://www.ft.com/content/59949c24-0a3a-11e3-9cec-00144feabdc0>
- Xu, S. (2009). *Development and Test of Dynamic Congestion Pricing Model*. Massachusetts, USA: Massachusetts Institute of Technology.
- Yildirim, M. B. (2001). *Congestion Toll Pricing Models and Methods for Variable Demand Networks*. Miami, FL, USA: University of Florida.
- Yu, B., Zhang, L., Guan, F., Peng, Z., & Yao, B. (2017). Equity based congestion pricing: considering the constraint of alternative path. *Operational Research*, 17(1), 313-337.
- Yu, H., & Zhang, M. (2017). Data pricing strategy based on data quality. *Computers & Industrial Engineering*, 112(October 2017), 1-10.
- Zeyu, J., Shuiping, Y., Mingduan, Z., Yongqiang, C., & Yi, L. (2017). Model Study for Intelligent Transportation System with Big Data. *Procedia Computer Science*, 107, 418-426.
- Zhang, H., Zhang, Q., & Chen, W. (2019). Bi-level programming model of truck congestion pricing at container terminals. *Journal of Ambient Intelligence and Humanized Computing*, 10(1), 385-394.
- Zhong, R. Y., Newman, S. T., Huang, G. Q., & Lan, S. (2016). Big Data for supply chain management in the service and manufacturing sectors: Challenges, opportunities, and future perspectives. *Computers & Industrial Engineering*, 101(November 2016), 572-591.
- Zhu, L., Yu, F. R., Wang, Y., Ning, B., & Tang, T. (2019). Big Data Analytics in Intelligent Transportation Systems: A Survey. *IEEE Transactions on Intelligent Transportation Systems*, 20(1), 383-398.